

# 質問応答に基づく対災害情報分析システム

後藤 淳<sup>†</sup>・大竹 清敬<sup>†</sup>・Stijn De Saeger<sup>†</sup>・橋本 力<sup>†</sup>・  
Julien Kloetzer<sup>†</sup>・川田 拓也<sup>†</sup>・鳥澤健太郎<sup>†</sup>

本論文では、地震や津波などの災害時に個人からソーシャルメディア上に発信される大量の書き込みから、救援者や被災者が欲している情報を自動的に取得する情報分析システムについて報告する。このシステムでは、質問応答技術により、災害時の被災地の状況や救援状況を俯瞰的に把握し、被災地からの想定外も含めた情報を取得することを目的としている。システムで利用している質問応答処理では構文パターンの含意に基づき質問文を拡張し、ソーシャルメディアへの書き込みに対して地名・場所名を補完することにより、幅広い質問に対応する。さらに、本システムを拡張することにより、被災地からの重要な情報提供が必ずしも救援者へ届かない問題に対応できることについて述べる。NPOや自治体などの救援者が状況把握のための質問を予め登録しておけば、救援を望む被災者がTwitterやBBS等へ書き込んだ時点で、情報を求める側と提供する側の双方に自動的に通知できる。これにより救援者と被災者の双方向のコミュニケーションが担保され、救援活動がより効率的になると期待される。本システムの質問応答性能を我々が用意した300問のテストセットのうち回答が対象データに含まれる192問を用いて評価したところ1質問あたり平均605.8個の回答が得られ、再現率は0.519、適合率は0.608であった。

キーワード：情報システム、質問応答、災害、ビッグデータ

## A Disaster Information Analysis System Based on Question Answering

JUN GOTO<sup>†</sup>, KIYONORI OHTAKE<sup>†</sup>, STIJN DE SAEGER<sup>†</sup>, CHIKARA HASHIMOTO<sup>†</sup>,  
JULIEN KLOETZER<sup>†</sup>, TAKUYA KAWADA<sup>†</sup> and KENTARO TORISAWA<sup>†</sup>

In this paper we introduce an information analysis system that automatically acquires from social media like Twitter the kind of vital information that rescue workers or disaster victims need in case of large-scale disasters like earthquakes or tsunamis. This system uses question-answering (QA) technology with the aim of helping users get a comprehensive overview of the state of affairs and the various conditions of ongoing rescue efforts in the afflicted areas, and detect potentially unanticipated information from the disaster areas. The system expands the input question with various entailment expressions, and augments the original social media input data by analyzing the mentioned place names in order to handle a wide variety of questions relevant to disaster scenarios. Moreover, we extend this system to address rescue workers' crucial problem of getting access to relevant information from the disaster areas in

---

<sup>†</sup> 情報通信研究機構, National Institute of Information and Communications Technology

times of crisis. To tackle this problem we introduce a mechanism by which rescue workers from NPOs or municipalities can register certain questions for situation assessment in advance, so that when a disaster victim posts some urgent requests for food, medicines or other essentials on Twitter or some other BBS, both information sender and information requester are automatically notified of this. We expect that such a mechanism can safeguard the two-way communication between rescue workers and disaster victims, and ultimately lead to a more effective rescue effort. We evaluate the system on a test set of 300 questions and their answers. For 192 questions whose answers are actually included in our system's index, we obtained on average 605.8 answers per question, with 51.9% recall and 60.8% precision.

**Key Words:** *Information Analysis, Question Answering, Disaster, Big Data*

## 1 はじめに

2011年3月11日に起こった東日本大震災では、テレビ、ラジオなどの既存メディアが伝えられなかった局所的な情報を、Twitterなどの個人が情報発信できるソーシャルメディアが補完する可能性を改めて知ることとなった。一方で、Twitter等で発信された大量の情報を効率的に把握する手段がなかったために、被災地からの切実な要望や貴重な情報が、政府、地方自治体、NPOなどの救援団体に必ずしも届かず、救援活動や復興支援が最大限の効率で進展しなかったという可能性も高い。我々が震災時のTwitterへの書き込み (tweet) を調査したところ、少なくとも救援者が何らかの対応をしたことを示す tweet が存在しない要請 tweet も非常に多く存在した。さらには大量に飛び交うデマを含む情報に振り回された人も多く出た。

こうした状況に対応するため、自然言語処理を用いてTwitter上の安否情報を整理することを目指した「ANPLNLP」の取り組みが行われたが、開発の速度や多数のボランティアを組織化するには課題があったことが報告されている (Neubig, Matsubayashi, Hagiwara, and Murakami 2011)。実際に災害が発生してから、新たにTwitter等のソーシャルメディアに自然言語処理を適用し、情報を整理する技術を開発するのは非常に困難であろう。我々は、将来起きる災害に備えて、事前にそうした技術を開発しておくことが極めて重要であると考えている。

また、我々が被災地で行ったヒアリングでは、現地からの要望とその支援とのミスマッチも明らかになっている。例えば、テレビや新聞などのマスメディアで伝えられた「被災地で防寒着が不足している」という情報に呼応して、多くの善意の人から防寒着の上着が大量に現地に送られたが、津波被害にあい泥水の中で復旧作業をする必要のあった人々がより切実に求めているのは、防寒のズボンであった。別の例では、全国から支援物資として届けられた多くの衣類はどれも通常サイズのものばかりで、4Lサイズなどの大きな衣類が必要な人が一月以上も被災時の衣類を着続ける必要があった。これらは、大規模災害発生時に生じる被災者の要望の広範さや事前にそうした要望を予測しておくことの困難を示す事例と言えよう。さらに、本論文

で提案するシステムで実際に tweet を分析したところ、被災地で不足しているものとして、「透析用器具」「向精神薬」「手話通訳」など平時ではなかなか予想が困難な物資が実際に不足している物品として tweet されていることも判明している。こうしたいわば想定外の要望を拾い上げることができなれば、再度要望と支援のミスマッチを招くこととなる。

以上が示唆することは、次回の大規模災害に備えて、ソーシャルメディア上の大量の情報を整理し、上述した想定外の要望も含めて、必要な情報を必要な人に把握が容易なフォーマットで届ける技術の開発を災害発生以前に行っておくことの重要性である。また、我々が備えるべき次の災害が、今回の震災と類似している保証はない。以上のような点に鑑みて、我々は想定外の質問も含め、多様な質問に対して、ソーシャルメディア上に書き込まれた膨大な情報から抽出された回答のリストを提示し、状況の俯瞰的把握を助けることができる質問応答システムが、災害時に有効であると考えている。

ここで言う俯瞰的把握とは、災害時に発生する様々な事象に関して、それらを地理的、時間的、意味的観点から分類した上でそれらの全体像を把握することを言う。別の言い方をすれば、その事象がどのような地理的、時間的位置において発生しているのか、あるいはそもそもその事象がどのような事象であるのか、つまりどのような意味を持つ事象であるのか、等々の観点でそれら事象を分類し、また、それらを可能な限り網羅的、全体的に眺めわたし、把握することである。このような俯瞰的把握によって、救援者サイドは、例えば、重大な被害が生じているにもかかわらず、炊き出し、救援物資の送付等が行われていないように見える地点を割り出し、なんらかの齟齬の確認や、救援チームの優先的割当を行うことが可能になる。あるいは各地において不足している物資を、例えば医薬品、衣類、食料といった観点で整理して、救援物資のロジスティクスを最適化するなどの処置も可能になる。さらに、こうした俯瞰的把握によって、上で述べたような想定外の事象の発見も可能になり、また、それらへの対処も容易になろう。逆に言えば、誰かがこうした俯瞰的把握をしていない限り、各種の救援活動は泥縄にならざるを得ず、また、想定外の事象に対してはシステムティックな対応をすることも困難となる。

また、被災者自身も現在自分がいる地点の周辺で何がおきているか、あるいは周辺にどのようなリソースが存在し、また、存在しないかを全体として把握することにより、現地点にとどまるべきか、それとも思い切って遠くまで避難するか判断が容易になる。避難に至るほど深刻な状況でなかったとしても、周辺地域での物資、サービスの提供の様子を全体として把握することで、物資、サービスを求めて短期的な探索を行うか否かの決断も容易になろう。我々の最終的な目標は、多様な質問に回答できるような質問応答システムを開発することによって、災害時に発生する tweet 等のテキストデータが人手での処理が不可能な量となっても、そこに現れる多様で大量の事象を意味的観点から分類、抽出可能にし、さらに回答の地図上への表示や、回答に時間的な制約をかけることのできるインターフェースも合わせて提供することによ

り，以上のような俯瞰的把握を容易にすることである。

本論文では，以上のような考察に基づき，質問応答を利用して，災害時に個人から発信される大量の情報，特に救済者や被災者が欲している情報を tweet から取得し，それらの人々の状況の俯瞰的把握を助ける対災害情報分析システムを提案する．将来的には本システムを一般公開し，被災地の状況や救済状況を俯瞰的に把握し，被災地からの想定外の要望をも取得し，効率的な救済活動につなげることを目指す．本論文では提案したシステムを実際に東日本大震災時に発信された tweet に適用した評価実験の結果を示すが，この評価においては以上のような被災状況の俯瞰的把握を助ける能力を評価するため，質問応答の再現率に重点をおいた評価を行う．逆に言えば，いたずらに回答の上位の適合率を追うことはせず，再現率の比較的高いところでの評価に集中する．また，本システムを拡張することで，被災者と救済者の間でより適切な双方向のコミュニケーションが実現可能であることも示す．こうした双方向のコミュニケーションはより適切かつ効率的な救済活動のために極めて重要であると考えている．

本論文で提案するようなシステムは非常に多くのモジュールからなり，その新規性を簡潔にまとめることは難しいが，本論文においては以下の手法・技術に関して我々のタスクにおける評価，検証を行った．特に C については，新規な技術であると考えている．

**A** 固有表現認識 (NER) の有効性

**B** 教師有り学習を用いた回答のランキング

**C** 含意関係認識における活性・不活性極性 (Hashimoto, Torisawa, De Saeger, Oh, and Kazama 2012) の有用性

ここで，A，B に関しては本論文における実験の目標ならびに設定では有効性は認められず，最終的なシステムではこれらの技術を採用しなかった．これらに関して現時点での我々の結論は以下の通りである．NER はそれ単体では，我々のタスクでは有効ではなく，その後の処理やそこで用いられる辞書等との整合性がとれて初めて有効になる可能性がある．また，回答のランキングは，我々の目標，つまり，少数の回答だけではなく，想定外も含めた回答を可能な限り網羅的に高精度で抽出することには少なくとも現状利用可能な量の学習データ，素性等では有効ではなかった．一方で，含意関係認識において活性・不活性極性を利用した場合，再現率が 50 % 程度のレベルにおいて，適合率が 7 % 程度上昇し，顕著な性能向上が見られたことから，提案手法にこれを含めている．

本論文の構成は以下の通りである．まず，2 節において本論文で提案する対災害情報分析システムの構成とその中で使われている質問応答技術について述べる．3 節では，人手で作成した質問応答の正解データを用いたシステムの評価について報告する．4 節にて上述した双方向のコミュニケーションの実現も含めて今後の本研究の展望を示す．さらに 5 節にて関連研究をまとめ，最後に 6 節にて本論文の結論を述べる．

## 2 質問応答に基づく対災害情報分析システム

### 2.1 システム構成

本システムは,「宮城県で孤立しているのはどこですか」,「福島県で何が不足しているか」など,自然言語の質問を入力とし,大規模な tweet コーパスからその回答と思われる表現を抽出し,ユーザに提示する。(なお,現在,システムは Twitter を主たる情報源としているが,掲示板や一般の Web 文書などにももちろん適用可能である。)図 1 に示すように,システムは tweet から構文パターンを抽出しインデックスを作成する回答インデックス作成モジュールと,回答検索時に使用する含意パターンデータベースを作成する含意パターン獲得モジュール,作成されたインデックスを用いて回答を抽出する質問応答モジュール,ユーザから入力された質問に対する大量の回答を効果的に提示する入出力モジュールから成る。

各モジュールの動作の概要は次の通りである。回答インデックス作成モジュールでは,まず tweet を文単位で形態素解析,構文解析し,地名補完モジュールにて処理された構文解析結果から,詳細については後述するパターンや周辺名詞句を抽出し,これを回答インデックスに含める。

含意パターン獲得モジュールは,大規模な Web コーパスを形態素解析,構文解析したデータから,含意関係にあるパターン(例えば,「X から Y まで歩く」は「X から Y まで移動する」を含意する)を自動的に抽出し,含意パターンデータベースを作成する。

質問応答モジュールは,ユーザから入力された質問をインデックス作成モジュールと同様に形態素解析,構文解析を行い,質問文からパターンや周辺名詞句を取得する。次に,質問文に含まれるパターンを用いて,含意パターンデータベースを参照し,最大で数千個程度の含意パ

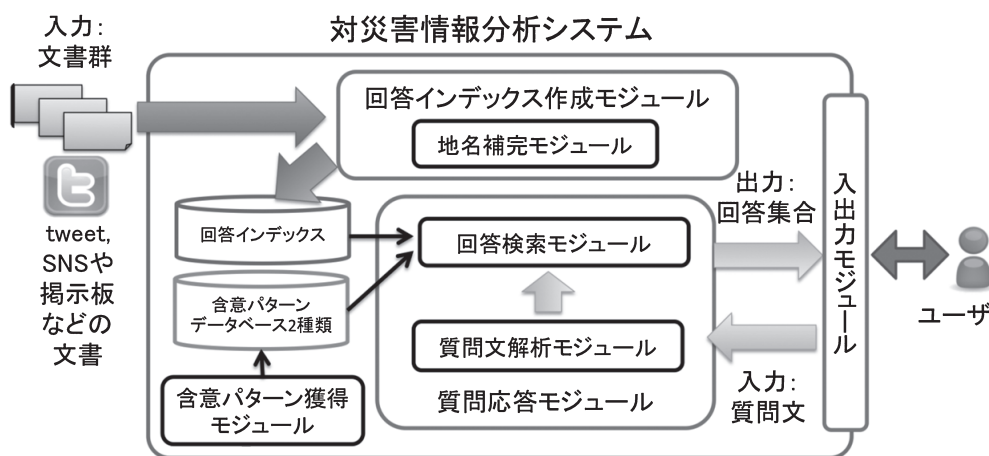


図 1 対災害情報分析システムの概要

ターンに拡張する。拡張されたパターンや周辺名詞句を用いて回答インデックスを検索し、回答を得る。

入出力モジュールは、2種類ある表示モードの選択、質問文の入力フォームなどを備え、ユーザーから入力があるとそれを質問応答モジュールに渡す。質問応答モジュールから回答を受けると、表示モードに応じてユーザに回答を表示する。

以下では、これらのモジュールの各々について説明する。

## 2.2 回答インデックス作成モジュール

回答インデックス作成モジュールは、大規模な tweet のデータを対象に、高速に質問応答を行うためのインデックスを作成するモジュールである。回答インデックスの作成には、Apache Jakarta Project のもとで開発が進められている Lucene<sup>1</sup>を利用する。以下ではこのインデックスを回答インデックスと呼び、その役割と作成手順、作成に際して注意が必要な地名の補完処理について説明する。

### 2.2.1 回答インデックスの作成

回答インデックスは、ユーザーから入力された質問文から生成したクエリを用いて高速に回答を取得するためのインデックスである。回答インデックスには、構文情報が十分に存在する文から抽出される情報を格納する回答インデックス 1 と構文情報が充分にない文から抽出される情報も格納の対象とする回答インデックス 2 の 2 種類がある。

回答インデックスの作成手順として、まず、対象 (tweet) を文単位で形態素解析、構文解析処理を行う。形態素解析には MeCab<sup>2</sup>、構文解析には日本語係り受け解析器 J.DepP<sup>3</sup> を使用する。

次に、回答インデックス 1 に格納するデータを作成するために構文解析結果における任意の名詞句 2 つとそれらをつなぐ文節係り受けのパスを構成する表層上の連鎖を取得する。例えば、「[宮城県の][炊き出し]」からは、「宮城県」と「炊き出し」という名詞句に係り受けのパスがあるので「宮城県の炊き出し」が取得される。一方、「[宮城県で][炊き出しが][行われる]」という結果からは、「宮城県」と「炊き出し」という名詞句の間に「行われる」という文節で媒介されるパスが存在するので「宮城県で炊き出しが行われる」が取得される。このパスを構成する 2 つの名詞句それぞれを変数で置き換えたものを構文パターン、あるいはパターンと呼び、また構文パターンとそれに含まれる変数に対応する名詞句 2 つの三つ組みをパターントリプルと呼ぶ。上記の「宮城県で炊き出しが行われる」という文からは、構文パターンとして「X で Y が行われる」、変数 X, Y に対応する 2 つの名詞句として「宮城県」と「炊き出し」の三つ組みが

<sup>1</sup> <http://lucene.apache.org/core/>

<sup>2</sup> <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html> 辞書は JUMAN 体系のものを使用。

<sup>3</sup> <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/jdepp>

この文から抽出されるパターントリプルとなる。またパターントリプルを含む tweet 内の名詞句を全てを周辺名詞句として取得する。最終的に、回答インデックス 1 には、パターンとして「X で Y が行われる」、変数に対応する名詞句としてそれぞれ「宮城県」「炊き出し」がキーに登録され、その値には変数に対応する名詞句と当該 tweet の ID が格納される。

回答インデックス 2 は、回答インデックス 1 に比べて、構文情報が不十分な文も対象とするために用いる。したがってこのインデックスを用いた回答の信頼性は高くないが、より広範な回答を得るために使用する。このインデックスでは構文パターンのかわりに部分パターンと呼ばれるパターンとその周辺名詞句をキーとする。構文パターンは、構文解析結果において二つの名詞句をつなぐパスから作られたが、部分パターンは名詞句一つと動詞、名詞、形容詞のいずれかへの係り受け関係から作られる。例えば、「宮城県です。透析用器具が足りません。」といった tweet からは任意の名詞句 2 つの間に係り受けが存在しないため、構文パターンを抽出することはできない。したがって、「透析用器具が足りない」という情報は回答インデックス 1 には反映されない。そこで、構文解析結果において「透析用器具」が助詞「が」を介して「足りません」へ係っているのを、係り元の名詞句を変数として部分パターンを抽出する。この場合は「X (= 透析用器具) が足りません」が抽出され、それと回答インデックス 1 同様に周辺名詞句である「宮城県」「状況」「透析用器具」とをキーとして、変数に対応する名詞句、すなわち「透析用器具」と tweet の ID とを値として回答インデックス 2 に登録する。

以上 2 種類の回答インデックスのキーと値を表 1 にまとめる。回答インデックス 1 は上述したパターントリプルを用いて作成したインデックスであり、回答インデックス 2 は、パターントリプルが取得できない tweet にも対応することで、更に幅広い回答を取得するためのインデックスである。以下に、回答インデックスを用いて、どのように回答を取得するかを説明する。

回答インデックス 1 では、例えば、「震災後、宮城県で透析用器具が不足しています」という tweet からは、パターンとして「X で Y が不足しています」、名詞句対 (名詞句 1, 名詞句 2) としてそれぞれ「宮城県」「透析用器具」、周辺名詞句として「震災後」「宮城県」「透析用器具」「不足」がキーに登録され、その値には変数に対応する名詞句と当該 tweet の ID が格納される。このようなエントリは、例えば「宮城県で何が不足していますか?」といった質問の回答を取得する際に使われる。この場合、インデックス検索時のクエリは「X で Y が不足しています」

表 1 回答インデックス

インデックス名	キー	値
回答インデックス 1	パターン, 名詞句 1, 名詞句 2, 周辺名詞句	回答 (= 名詞句 1, 名詞句 2), tweetID
回答インデックス 2 (部分パターン用)	パターン, 周辺名詞句	回答 (= 名詞句 1), tweetID

というパターンと、「宮城県」という名詞句 1 であり、検索の結果、上述した tweet の例から生成されるインデックスのエントリに値として登録されている名詞句 2 の「透析用器具」が回答として、tweet の ID とともに出力される。また、「どこで透析器具が不足していますか？」という質問であった場合には、「X で Y が不足しています」というパターンと「透析用器具」という名詞句 2 を持つクエリが生成され、値に登録されている名詞句 1 の「宮城県」が回答として、tweet の ID とともに出力される。なお、上では周辺名詞句がキーとして登録されると説明したが、Lucene のインデックスのメカニズムでは、キーの一部を省略することが可能であり、例えば、上の質問の例では、パターントリプルを抽出してきた tweet にあった「震災後」という名詞句はクエリ中のキーとして現れないが、適切に検索が行われる。

一方、回答インデックス 2 のエントリは、例えば、「宮城県で何が足りませんか？」という質問に対する回答を得るためにも使うことができる。質問中では、「宮城県」は「足りません」という動詞にかかっているが、この宮城県を周辺名詞句としてとらえ直し（回答が含まれる tweet として「宮城県です。～が足りません」のようなものもあると想定する）、「何が足りませんか」という質問中の部分から「X が足りません」という部分パターンを作成すると回答インデックス 2 を検索できる。本来であれば、先の tweet の解析時に照応解析等を行い、「透析器具が足りません」という文には「宮城県で」という表現が省略されていることを認識した上で処理を進めるべきであるが、そもそも照応解析等の精度が高くない現状に鑑み、照応、省略表現を一括して周辺名詞句として扱うことで柔軟な回答の抽出を狙っていることになる。

なお、いずれのインデックスの作成時においても、retweet が入力として与えられた場合には、同一内容の retweet があるかをチェックし、もし存在すれば 1 つの retweet のみを登録し、これと同一内容の複数ある retweet はインデックスには登録しない。一方ですべての retweet の ID のリストは別途保存しておく。これは retweet の処理によって質問応答の処理時間がのびるのを防ぐための処理である。

### 2.2.2 地名補完モジュール

地名補完モジュールは、回答インデックスの作成の際に、tweet などのソーシャルメディアへの書き込みで省略されがちな地名や場所名を補完するモジュールである。地名補完モジュールでは大きく分けて次の二つの処理を行う。(1) まず、構文解析結果をその入力とし、地名補完の対象となるエンティティを認識する。(2) 認識されたエンティティの詳細な住所情報を取得し、元のエンティティの周辺情報に基づいて後述する場所の包含性や、場所の非明示性の問題に対処する補完処理を行い構文木に適宜補完要素を挿入する。

災害に関する情報では、効率的な救援活動などのため、位置情報や地名が極めて重要である。Twitter では、携帯端末等 GPS 情報を付加できる装置からの書き込みの場合、位置情報の開示設定がされていれば、その tweet が書き込まれた場所を特定できる。しかしながら、多くのユー



ずは、プライバシー等の問題から該当機能を有効にはしていない。災害時の要望等については、この機能を有効とすべきであるが、かならずしもすべての情報に位置情報が記述されている訳ではない。さらに、通信が不可能なほど壊滅的な被害が発生した場所から、通信が可能な地域に移動し、当該地域について tweet する場合など、tweet がなされる位置とその tweet が言及している位置が、一致しない場合もある。そのため、tweet 内の地名を特定し、適切に処理することが重要である。しかしながら、地名の処理には以下のような問題があり、極めて難しい課題となっている。

**場所の非明示性：** Twitter などへの書き込みには、明示的に県や市の名称が書かれていないことが多い。さらには、tweet に限らず、一般的に、イベントが起きた場所を指す名詞句がイベントを表す動詞等に明示的には係らないことも多く、動詞で表されたイベントと地名を結びつけることはそれほど容易ではない。

**場所の包含性：** 場所には包含性がある。例えば、仙台市が宮城県の中にあることを正しく認識しても、それを処理する手だてがなければ、たとえ文中に「仙台市」と記述されていても、「宮城県で」と問う質問には回答できないということが起きる。

**場所の曖昧性：** 一部の地名は非常に大きな曖昧性を持ち、上記の包含性を扱おうとする場合に、特に問題となる。例えば、「福島」という地名は日本全国に 50 以上もあり、そこから正しい一つを選ぶ必要がある。

地名補完モジュールにて解決したい問題とほぼ同一の問題に取り組んでいるプロジェクトとして GeoNLP<sup>4</sup> がある。また、地名をはじめとする固有表現の認識という点では、近年 Twitter 等のソーシャルメディアに対する固有表現認識の難しさや、問題点が広く知られ、報告も多くなりつつある (Liu, Wei, Zhang, and Zhou 2013; Ritter, Clark, Mausam, and Etzioni 2011; Cheng, Caverlee, and Lee 2010)。Liu らは tweet を対象として K-Nearest Neighbors と Conditional Random Fields を組み合わせた新しい固有表現認識器を提案している。Ritter らは Labeled LDA に distant supervision を適用することで高い性能を持つ固有表現認識器を実現している。また、Cheng らは、tweet のみならず Web コーパスを用いた教師なし学習による固有表現認識器を提案している。

前述した問題に完全に対応することは難しいが、現在のシステムは以下の手続きによって、地名とイベントとを対応付けている。具体的には、まず、現在入手可能なデータから大規模な地名・場所名辞書を自動生成し、さらに、地名等の包含性、曖昧性の一部をヒューリスティクスによって対処しつつ、回答インデックスに地名の情報を取り込んでいる。以下ではこの各々のステップについて説明する。

---

<sup>4</sup> <http://agora.ex.nii.ac.jp/GeoNLP/>

### 2.2.3 地名・場所名辞書の作成

地名補完の対象となるエンティティを特定するため、日本郵便が公開している郵便番号データと Wikipedia に基づく上位下位関係 (Yamada, Torisawa, Kazama, Kuroda, Murata, De Saeger, Bond, and Sumida 2009) を利用して、地名・場所名辞書を作成した。

まず、日本郵便が公開している郵便番号データを用いて地名辞書を作成した。郵便番号データからは、「都道府県／市区町村／町域」で表される住所の情報から、用いられる可能性がある地名文字列とその詳細な住所との対応を取り出す。地名文字列は「山元」のように断片的なものである場合が多いが、こうした対応づけを用いて、断片的な文字列から「宮城県亘理郡山元町」のようなより詳細な住所が入手可能となる。さらに、「都道府県／市区町村／町域」という住所の階層性は、先に挙げた場所の包含性に対処するための情報源となる。このようにして、2,486,545 のエントリを持つ辞書（地名辞書）を作成した（地名文字列—住所の対の数は 5,129,162）。そのうち、84,633 エントリが曖昧性をもつ地名であった。

また、Twitter などへの書き込みでは、住所のような地名の他に学校や施設、ランドマーク的名称の正式名称から通称までが幅広く用いられる。そこで、Wikipedia から抽出した上位下位関係 (Yamada et al. 2009) から、上位語として自治体を取り、「(自治体名) の (\*X)」(X は「施設」「学校」など) というパターンにマッチする下位語を取り出して利用した。例えば、「名取市の増田小学校」などである。これは、「学校」などの、郵便番号データには載っていないような場所にもその詳細な住所を対応づけるためである。上位語中の自治体名を、地名辞書で検索して下位語に住所を付与する。最終的に、255,273 エントリを持つ場所辞書を作成した。

地名辞書と場所辞書をマージすることで、2,741,818 エントリを持つ辞書が得られる。地名辞書も場所辞書もほぼ全自動で作成しているため、それをそのまま文字列マッチによる単純な地名検出手法とともに適用した場合には、問題となる場合がある。例えば、「枝野官房長官」の名字と同じ「枝野」が宮城県の地名として使われている場合があるなど、地名には人名と同じものが多くあり周辺の情報から適切に処理される必要がある。また、高頻出な普通名詞をいずれかの辞書のエントリとして含んでおり、誤って地名処理される場合もある。そこで、このような問題となるエントリを可能な限りマージした辞書から人手で取り除いた。その結果、2,726,944 エントリを持つ地名・場所名辞書が得られた。

地名・場所名辞書は、地名補完モジュールの性能を決定する極めて重要な知識である。人工物に対する固有表現ほど新規エントリや、変更があるとは考えていないが継続的にメンテナンスされる必要がある。このような知識は、ひとたび整備されれば、その多くは長期にわたって利用可能であるためコストをかけ整備する価値があると考ええる。

#### 2.2.4 地名・場所名特定

回答インデックスを作成するために形態素解析, 構文解析がされた解析結果の各文節に対し, 形態素をその単位として最長の名詞句を抽出し, 地名・場所名辞書を用いて地名・場所名を特定し, 当該名詞句に詳細な住所候補を付与する. その際, 名詞句全体がマッチしない場合でも, その範囲内で最左のマッチを選び, できるだけ住所を付与する. なお, 1 文字の地名・場所名は誤ったマッチである可能性が大きいため, 無視する.

現在のシステムの地名・場所名の特定方法は, 形態素を単位とする表層文字列が地名・場所名辞書に存在するか否かによって行うため, 一般名詞等を誤って地名・場所名として扱う場合がある. そこで, 地名・場所名の特定に関して, 通常の固有表現認識器を用いることが考えられる. 風間らの報告 (風間, De Saeger, 鳥澤, 後藤, Varga 2012) では, 固有表現認識器の有効性が確認されておらず, 我々の実験においてもその有効性を確認できなかったため, 現在のシステムでは, 固有表現認識器を用いていない. 実験の詳細については, 3 節にて述べる.

上記の問題以外にも, 本システムでは, 情報が無ければ最も広範囲な地域を表す住所, 直前に曖昧性解消された住所がある場合には, それと最も整合性のある住所を選ぶルールに基づく曖昧性の解消を行っている. 候補のうち, 県・郡・市 (郡部の場合は町) 部分が tweet 中の文字列と一致すれば, より広い地域レベルで文字列と一致しているものを優先する. 例えば, 「福島」の場合には, 「福島県: 福島市」, 「大阪府: 大阪市: 福島区」等数多くの曖昧性があるが, 最も広範囲な「福島県」が選択される.

#### 2.2.5 地名補完処理と回答インデックスへの反映

本システムでは, 「イベントの場所は文中で直前に出現した地名・場所である」という仮定を置き, 元の文の構文解析結果を操作し, 直前の地名・場所 (tweet が複数文の場合は前方の文も考慮する) に場所を表す助詞「で」を加えたものを, イベントを表す動詞等に係るように付け加えた新たな構文解析 (補完構文解析) 結果を生成する.

例えば, 「気仙沼中学校へ避難しています」という文があった場合, 「避難」イベントの場所は, 直前の場所である「気仙沼中学校」と認識され, さらに地名・場所辞書により「気仙沼中学校 → 宮城県 / 気仙沼市」であると分かっているとすると, 「宮城県で」, 「気仙沼市で」などの助詞「で」で終わる複数の文節が元の構文木に挿入される. こうしてできた補完構文解析結果を利用することで, 補完された場所に関連する質問に対応したインデックスが生成される. これにより, 例えば, 元の文には「宮城県」という表現が含まれていないにもかかわらず「宮城県でどこへ避難していますか」という質問に対し回答 (= 気仙沼中学校) できる.

## 2.3 含意パターン獲得モジュール

含意パターン獲得モジュールでは、大規模なコーパスから含意パターンを獲得し、それをデータベース化する。含意パターンとは、簡単に言うと、あるパターン「X から Y まで移動する」を含意する「X から Y まで歩く」のようなパターンのことであるが、含意が成立するための名詞句 X, Y にある制約等を考慮するといくつか種類が考えられる。ここでは、クラス依存のパターン、クラス非依存のパターンと部分パターンという三種類の構文パターンの含意パターン獲得及びそのデータベース化について説明する。

### 2.3.1 クラス依存のパターン

クラス依存パターンとは、パターン中の変数に対応する名詞の意味クラスに制約を掛けた構文パターンである。構文パターンにクラス制約を掛けることでパターンの多義性が解消できる。例えば、「Y のための X」という構文パターンは「Y:病名のための X:薬品」のように、Y が病名、X が薬品の意味クラスの単語の場合は、X と Y の治療関係とでも呼べる関係を表し、上記のパターン「X:薬品で Y:病名が治る」の含意パターンとみなせるであろう。一方、「X:作業のための Y:道具」の場合は手段または道具という意味的關係を表現する。このようにして構文パターンと共起する単語を特定の意味クラスに限定することで、構文パターンの曖昧性が大きく減らされ、高頻度で曖昧なパターンが活用可能になり、より大量の回答を獲得できる (De Saeger, Torisawa, Kazama, Kuroda, and Murata 2009)。

意味クラスは、Kazama ら (Kazama and Torisawa 2008) が提案した単語クラスタリング法によって自動獲得する。この手法では大規模 Web コーパスから得られる名詞と動詞の係り受け関係の統計データを用いて、名詞の隠れクラスへの事後確率の分布を求める。ある名詞の所属確率が 0.2 以上の隠れクラスを、その名詞の意味クラスとする。現状では名詞 100 万個を 500 クラスに分類したクラスタリングデータを用いる。

クラス依存の含意パターンの認識には Kloeitzer らが提案したクラス依存パターン間の教師付きの含意獲得手法 (Kloeitzer, De Saeger, Torisawa, Sano, Goto, Hashimoto, and Oh 2012) を用いる。詳細については (Kloeitzer et al. 2012) を参照されたいが、含意パターンを認識する SVM 分類器は主に次の 3 種類の手がかりを用いる。

- (1) パターンの表層的素性 (表層／構造を考慮した素性)。これらの素性は、表層上似ているパターンは含意関係にある可能性が高いという前提で、パターンに含まれる形態素、内容語、構文木の部分木などの bag of words 表現を基に計算した様々な類似尺度から成る。
- (2) 分布類似度に基づいた素性。ある構文パターンとその含意パターンの候補に関しては、6 億ページの日本語 Web 文書からパターンの変数に当てはまる名詞句対を検出し、それらの名詞句対の相対的なオーバーラップを計算する。例えば、「X で Y を提供」と「X で Y を配っている」という 2 つのパターンは X と Y の変数に頻出する共通の単語対 (例え

ば,「石巻市, 救援物資」が多ければ多いほど, これらの構文パターンがお互いの言い換え表現となっている可能性が高いと考えられる. 似た文脈に出現する語は似た意味をもつというのは, 分布仮説 (Harris 1954) と呼ばれる言語学におけるよく知られた仮説である. これらの素性はクラス依存のパターンの意味クラスに属する単語対に基づいて計算した類似尺度から成る.

- (3) 言語資源に基づいた素性. これらの素性は高度言語融合フォーラム ALAGIN で公開された動詞含意関係データベース (ALAGIN リソース A-2), 日本語異表記対データベース (ALAGIN リソース A-7), 基本的意味関係の事例ベース (ALAGIN リソース A-9) と日本語形態素解析器 JUMAN の辞書から得られた異表記と反対語データを言語資源として参照し, 両パターンに含まれる内容語が同義語あるいは異表記である場合, または含意関係や対義関係にある場合など, これらの言語資源に含まれる意味的關係にある時にその情報を素性に加える. 更に, Hashimoto らが提案した「活性・不活性テンプレート」(Hashimoto et al. 2012) も素性として用いる. この活性・不活性テンプレートについては後述する.

学習データは 51,900 サンプルであり, SVM での学習には 2 次の多項式カーネルを用いた. 図 2 は, 学習データとは異なる 5,338 の評価セットを用いて評価した本分類器から得られるクラス依存パターン含意の認識精度である.

図 2 から分かるように, 上述した条件ではこの手法の上位 1 億対 (データサンプル数 49) では約 85 % の適合率を示し, 上位 2.37 億にて約 70 % の適合率を保持している. 本論文のシステ

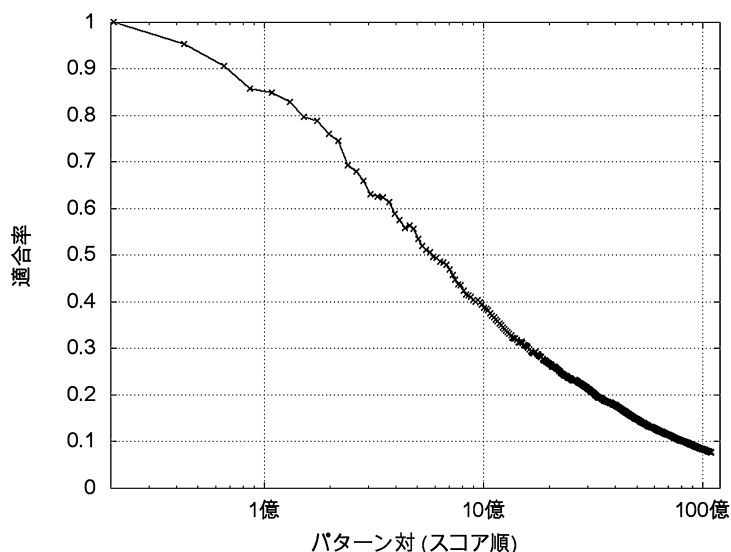


図 2 構文パターン間の含意認識の適合率

ムで利用される含意パターンデータベースは、後述する方法により質問文から得られる可能性のある構文パターンの含意パターンを SVM スコアが高いものにしばって格納しているので、回答検索に用いる含意パターンの適合率は図 2 に示される上位の適合率に相当するものと考えられる。

本システムで利用する含意パターンデータベースを構築するため、まず、(Kloetzer et al. 2012)と同様に、500 意味クラスの任意のペアのうちで、同じ名詞句対を異なり数で 3 つ以上共有するパターン対すべてを考える。こうしたパターン対の総数は 108 億個存在するが、そのすべてに対して、分類器を適用して SVM スコアを求める。ついで、SVM スコアが計算されたパターン対の内、以下の手続きで最終的な含意パターンデータベースを構築する。まず、上述のパターン対に含まれるパターンを「含意されるパターン」P として一つ選択し、SVM スコアが 0 以上のパターンを「含意するパターン」Q としてスコア上位から順に取得する。「含意するパターン」Q が 500 個を超えた場合は、スコア上位 500 個のみを「含意されるパターン」P と対にしてデータベースに格納する。この操作を 108 億個のパターン対に含まれるパターン各々を「含意されるパターン」P と仮定して繰り返す。なお、上位 500 個という数値は決定的なものではなく、システムのパラメータのひとつであるが、求める性能と応答速度のトレードオフによって決まる。現在の 500 という数値は、さまざまな質問をシステムに投入し、経験的に決めたものである。

### 2.3.2 クラス非依存パターン

クラス依存のパターンでは、特定の意味クラスの組み合わせにふさわしい含意表現を発見しやすい。一方、なるべく広い文脈で含意表現として通用するパターンも回答抽出に利用したい。そのために、入力パターンとそのクラス依存の言い換えパターンの集合をクラス非依存の含意パターン、つまり名詞句に何らの意味的制約が加えられていないパターンで補完する。多くの意味クラス対で含意パターンとして通用するものは恐らく非常にロバストで一般的な言い換え表現であるという前提を基に、クラス依存パターン間の各意味クラス対での SVM スコアを平均したパターン対のデータベースを用意する。あるパターンのクラス非依存の含意パターンは上記のクラス依存のケースと同様のアルゴリズムで選別する。例外処理として 1 つの意味クラス対としか共起しないパターンを除外する。

さらに、「Q が P を含意する」という関係におけるパターン Q とパターン P において、通常の「Q が P を含意する」場合のスコアと、逆向きの「P が Q を含意する」場合のスコアが両方向ともに 0 以上のパターン対のみに限定する。これは確かに片方向の論理的含意関係が成立しているものの、あまりに意味的にかけ離れているパターン対で回答を認識するのを防ぐためである。こうして集められた「含意するパターン」Q は、スコア上位 500 までの「含意されるパターン」P と共にデータベースに格納される。得られた Q が 500 個未満の場合には、その時点

までに登録されたすべての Q と同じ内容語（動詞，名詞または形容詞）を持つ P をスコアの高いものから順に取得し，データベースに登録する。

### 2.3.3 部分パターン

ソーシャルメディアから得られるテキストはインフォーマルな書き方で知られている。特に Twitter の場合では，tweet が 140 文字以内という制限があるので，必要最低限の情報しか含まない tweet が多い。そのため，二つの名詞句の存在を前提とするクラス依存パターンやクラス非依存パターンがうまく適用できない場合が非常に多い。この問題に対処するために上記のクラス非依存のパターンを一つの名詞句の存在を前提とする部分パターンに分割する。例えば，「X が Y で孤立する」という構文パターンはそれを構成する係り受け関係「X が孤立する」と「Y で孤立する」に分割される。

部分パターンの含意パターンデータベースを次のように用意する。既に説明したクラス非依存パターンの含意データベースを入力とし，それらのパターン対を分割し，変数毎に部分含意パターンの候補ペアを生成する。例えば，（「X が Y で孤立する」，「Y では X に連絡できない」）というクラス非依存パターン対から（「X が孤立する」，「X に連絡できない」）と（「Y で孤立する」，「Y では連絡できない」）という 2 つの部分パターン対を含意候補として生成する。この部分パターン対の含意スコアはクラス非依存の含意パターンと同様に，その生成元のクラス非依存の全含意パターン対のスコアの平均とする。ただし，生成元の含意パターン対が 1 つしかない部分含意パターンは一般性に欠けていると考え，除外する。さらに，クラス非依存パターンと同様に，「Q が P を含意する」と「P が Q を含意する」の両方向のスコアが 0 以上のパターン対のみをデータベースに登録する。

### 2.3.4 部分パターン対のクリーニング

以上の方法で作成した部分パターン対は，それがもたらされたクラス非依存パターン対のスコアを平均した値をスコアとして持っているが，パターンに含まれる用言相当表現と変数との関係を全く考慮していないため，信頼性を欠く場合がある。そこで，次の 2 つの方法で，部分パターン対をクリーニングする。

- 活性・不活性極性 (Hashimoto et al. 2012) を用いて部分パターン対を構成する 2 つのパターンの極性が異なる部分パターン対は削除する。
- 部分パターン対 (P-Q) においてパターンを構成する動詞が P と Q において同一であるが，変数とその動詞を媒介する助詞が異なる部分パターン対は削除する。例えば，「X が不足する」と「X に不足する」などの部分パターン対である。ただし，助詞「は」と「が」の組み合わせは許容し，削除しない。

ここで、活性・不活性極性とは、Hashimoto らが提案した新しい意味極性であり、助詞と動詞の組、すなわち本論文で言うところの部分パターンに対して活性、不活性、中立の 3 つの極性が付与されている。活性極性が付与された部分パターンはそれを埋める名詞の主たる機能、効果、目的、役割、影響が準備あるいは活性化することを意味し、その典型例としては「X を引き起こす」「X を使う」「X を買う」が挙げられる。不活性の部分パターンは逆にそれを埋める名詞の主たる機能、効果、目的、役割、影響が抑制あるいは不活性化されることを意味し、典型例は「X を防ぐ」「X が不足する」「X を破壊する」などが挙げられる。中立の部分パターンは活性、不活性のいずれも付与できない意味的性質を持つものである。

本研究で含意関係を持つものとして生成された部分パターン対には「X が不足する」「X が足りる」のように意味的には真逆であり、含意が成立していないものが多数含まれた。これは含意パターン認識で使われている分布類似度がこうした意味的差をとらえられないためであると考えられる。一方で、活性・不活性極性に従えば、「X が不足する」は不活性、「X が足りる」は活性であり、それらの差を見ることによって、意味的差異をとらえることができる。我々は、活性部分パターンを 11,276 個、不活性部分パターンを 2,764 個、中立部分パターン 7,523 個を人手でアノテーションしており、このデータを用いて、部分パターン対で極性が異なるものを削除した。

以上のクリーニングによって、当初 9,192,475 個の部分パターン対から 1,819,651 個の部分パターン対が削除され、最終的に 8,033,759 個の部分パターン対がデータベースに格納された。なお、このうち、活性・不活性極性によるフィルタリングの結果除かれた部分パターン対は 1,158,716 個であった。

## 2.4 質問応答モジュール

質問応答モジュールは、ユーザが入力した質問文から回答集合を出力するまでの一連のモジュールで構成される。具体的には、質問文から構文パターンを抽出する質問文解析モジュールと、インデックスから回答を検索する回答検索モジュールから構成される。以下に各々の説明を述べる。

### 2.4.1 質問文解析モジュール

質問文解析モジュールでは、自然言語で入力された質問文の格助詞の変更や疑問代名詞の位置の入れ替えなどをルールベースで行う。これは、複数の質問構文パターンを用いてより多くの含意パターンを獲得し、幅広い回答を取得するための処理である。次に、ルールベースで言い換えられた質問文の構文解析結果から疑問代名詞以外の名詞句一つと疑問代名詞を特定し、その間の係り受け関係パス上にある表現から構文パターンを取得する。例えば、「宮城県で何が不足していますか」という質問が入力された場合、「X (=宮城県) で Y (=何) が不足してい



る」という基本的な構文パターンに加え、「Y が X で不足している」(格要素の入れ替え), 「Y は X で不足している」「Y が X では不足している」「X で Y は不足している」「X では Y が不足している」(助詞の変換), 「X で不足している Y」(ガ格疑問代名詞の被連体修飾化)などの構文パターンが得られる。このようにして得られた構文パターンを用いて, 後述する回答検索モジュールで回答インデックスを検索するクエリが生成される。例えば, 「X (=宮城県) で Y (=何) が不足している」からは, パターンに「X で Y が不足している」, X に対応する名詞句 1 に「宮城県」を指定したクエリと, 部分パターンとして「Y が不足している」, 周辺名詞に「宮城県」を指定したクエリが得られる。

疑問代名詞以外に 2 つ以上の名詞句が含まれる場合は, 疑問代名詞と名詞句一つとそれをつなぐ文節で表される複数のパターンを抽出する。例えば, 「宮城県ではどこで携帯が充電できますか」が入力された場合, 「X (=宮城県) では Y (=どこ) で充電できる」, 「Y (=どこ) で X (=携帯) が充電できる」の構文パターンが取得される。この結果から, パターンに「X では Y で充電できる」, X に対応する名詞句 1 に「宮城県」, 周辺名詞句に「携帯」が指定されたクエリと, パターンに「Y で X が充電できる」, 名詞句 1 に「携帯」, 周辺名詞句に「宮城県」が指定されたクエリが生成される。同時に, 部分パターンとして「Y で充電できる」, 周辺名詞句に「宮城県」「携帯」が指定されたクエリも生成される。なお, クエリで指定される周辺名詞句は, 質問文に含まれる全名詞句から, パターンや名詞句 1 に含まれる名詞句を除外し作成される。

質問文解析モジュールでは, 質問構文パターンの獲得のほか, 疑問代名詞に助詞「は」とともに直接係る名詞がある場合, その名詞を主題語として取得する。例えば, 「被災地で不足している食べ物は何ですか」という質問が入力された場合, 名詞「食べ物」を主題語として取得する。この主題語は, 得られた回答との分布類似度 (Kazama and Torisawa 2008) により, 回答候補を選別するための情報として利用される。例えば, 「食べ物」に対して分布類似度が高い上位の名詞には, 「お菓子」, 「酒」, 「魚」, 「肉」, 「ワイン」, 「チョコレート」などの食べ物が含まれている。逆に食べ物と関連性の薄い「タオル」や「電化製品」の分布類似度は非常に低い。このように, 主題語と回答候補との分布類似度は, 質問の回答として相応しくない回答候補を除外する特徴として利用できる。

#### 2.4.2 回答検索モジュール

最終的な回答の取得に際しては, 質問文解析モジュールによって得られた複数の質問構文パターンから, 2.3 節で説明した含意パターンデータベースを引くことで質問構文パターンを含意する含意パターン集合が取得される。ついで, 質問構文パターンと質問文中で共起する疑問代名詞以外の名詞句と含意パターン, 質問文中の周辺名詞句などをキーとして回答インデックスが引かれ, 回答と回答が抽出された tweet の ID が得られる。

より具体的に述べると, 一つの質問から得られる複数個の質問構文パターンの各々につき,

最大で 1,500 個の質問構文パターンの含意パターンが生成される。その内訳はそれぞれデータベースに格納されているクラス依存パターンが最大で 500 個、クラス非依存パターンが最大で 500 個、部分パターンが最大で 500 個となる。これらのパターンは質問文中に出現する名詞句と組み合わせて回答インデックスの検索に使われる。また、各々の回答インデックスは本論文の実験では数千万件レベルの大量の tweet をカバーしているため、如何にこの回答インデックスを引く操作を高速化するかが重要になる。現在のシステムでは、Bloom Filter(Bloom 1970) を利用して、回答インデックスに共起がないパターンと名詞句の組み合わせから成るパターントリプルをメモリー上の操作のみで近似的に検出し、ディスクアクセスを伴う回答インデックスの検索回数を劇的に減らしており、これにより実用的な速度を得ている。

これまでも述べたとおり、二つの名詞句をつなぐ構文パターンと周辺名詞句をキーとする回答インデックス 1 は、質問文からパターントリプルが取得できた際に検索される。部分パターンをキーとする回答インデックス 2 は、二つの名詞句をつなぐ構文パターンが質問文から抽出されたときも含め、部分パターンが得られる場合すべてにおいて使用される。さらに、回答インデックス 2 に対して、パターンやその内容語を周辺名詞句として検索することで、パターンに直接係り受けない回答も取得できる。また、部分パターンに含まれる内容語のみをとりだし、それを周辺名詞句として検索することも行う。これは例えば「何が不足しているか？」という質問に対して、「不足」のみを周辺名詞句として検索することに相当する。

なお、抽出された回答にはストップワードフィルター、場所名フィルター、非場所名フィルターが適用される。ストップワードフィルターは、あらかじめ用意したストップワードリストに回答が含まれる場合にそれを回答リストから削除するものである。ここで使用しているストップワードリストは含意パターンデータベース構築の際に用いた 6 億ページの Web 文書から形態素態素解析器を使って自動的に認識された名詞句（複合語および単語）のうちで、明らかに解析ミスであり語として認められないものや非常に漠然としており明確な概念を指しているとは言えないもの（例：「皆さん」「双子以上」「その他」）、さらには主として機能語的に利用される語（例：「理由」「モノ」）を人手で集めたものである。これは現在 164,064 個の名詞句を含んでいる。

場所名フィルターは、疑問代名詞「どこ」を含む質問に関して、前述した地名・場所名辞書にある語を含む回答、前述した単語クラスタリングの結果から場所名をさす語を多く含む 48 クラスに含まれる語を含む回答、あらかじめ用意した「\*ホテル」、「\*センター」などの場所名のためのパターン 113 個に合致する回答のいずれでもないものを回答リストから削除する。一方で疑問代名詞「何」を含む質問に関しては、非場所名フィルターを適用する。これは場所名フィルターを逆に用いて地名フィルターでは削除される回答のみを最終的な回答リストに含めるフィルターである。

なお、回答が一文字の場合には、そもそも誤答である可能性が高く、また、後述する再現率

の計算において問題になるため、そもそも回答リストに含めないこととした。

## 2.5 入出力モジュール

入出力モジュールは、ユーザーから入力される質問を質問文解析モジュールに送信し、回答検索モジュールから出力される質問に対する複数の回答を提示する。本モジュールは Web ブラウザーを用いたインターフェースを備えており、一連の操作は Web ブラウザー上で操作できる。また、回答検索モジュールから出力される大量の回答の俯瞰的な把握を可能にするために、次に述べる 2 種類のモードで結果を表示する。ひとつは、回答結果を単語の意味クラス毎にまとめて表示するモードであり、もう一方は、場所を尋ねる質問に適した結果の表示方法として、地図上に回答を表示するモードである。以下で、それぞれについて説明する。

### 2.5.1 意味クラスを利用した回答表示モード（意味マップモード）

意味クラスを利用した回答表示モードでの実行例を図 3 に示す。この回答表示モードでは、回答が意味クラスごとにまとめられ、異なる色で表示される。色には意味はなく、異なる意味クラスクラスタであることを示すのみである。意味クラスは (Kazama and Torisawa 2008) で計算されたものを用いるが、意味クラスの計算対象外であるような長い名詞句に対しては、部分

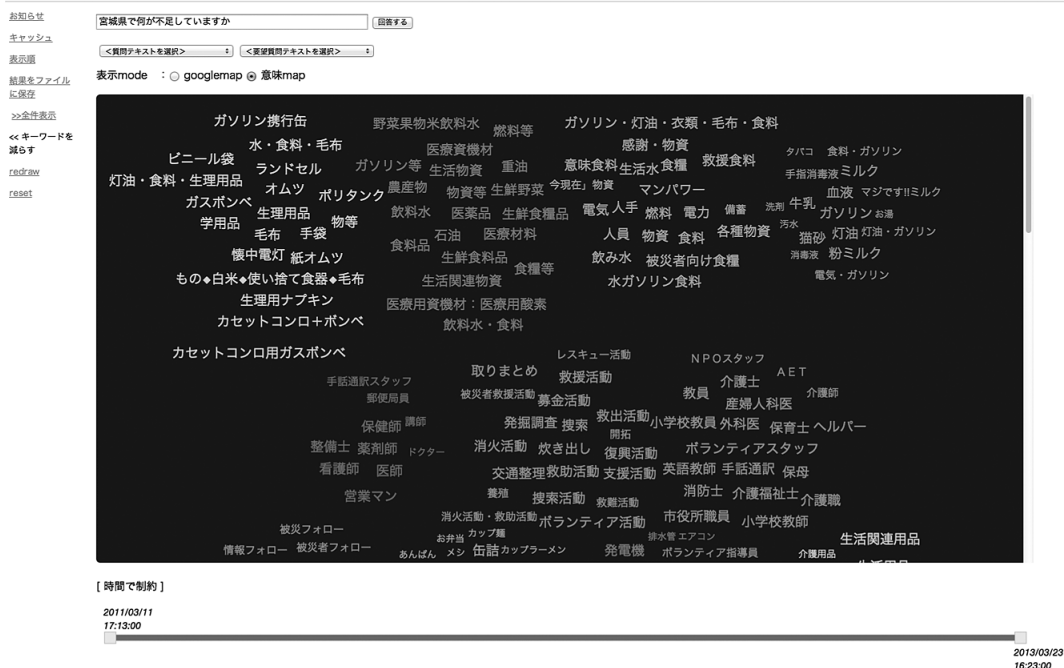


図 3 意味マップモードでの実行例

マッチを適用するなどして対応する。この表示方法によって、回答を俯瞰的に把握することが可能となる。回答の文字列をクリックすると、回答を抽出してきた情報源 (tweet) へのリンク、もしくは回答を抽出してきた tweet そのものの表示するウィンドウがポップアップし、回答が抽出された tweet の内容を確認できる。

また、画面下部にあるスライダーによって、情報抽出源のテキストの発信時刻による回答の限定が可能である。回答が抽出されたテキストの発信時刻は、一般の Web ページを対象とする場合は特定が困難であるが、Twitter や SNS (Social Networking Service) であれば、その情報を発信した時刻を容易に特定できる。スライダーによって時間帯を指定すると、その時間帯に発信されたテキストから抽出された回答のみが表示される。特定の期間に発信されたテキストからの回答が欲しい場合や、古くなった情報を非表示にしたい場合などには、この機能を用いて必要とする期間に回答をフィルタリングできる。

## 2.5.2 地図上へ回答を表示するモード (google マップモード)

回答を地図上へ表示するモードでの実行例を図4に示す。この表示方法では、質問の回答となる場所の位置が地図上で表示される。例えば、「宮城県のどこで炊き出しをしていますか」という質問に対して、炊き出しが行われている地点が容易に把握できるようになる。この表示モードにおいて、質問応答サーバーから受け取る情報は、意味マップモードの場合と同一である。このモードでは、地図上に回答を表示するために、次のことを行う。

- (1) 質問が場所を尋ねる質問（～はどこですか、どこで～できますかなど）の場合、回答は地名・場所名であることから、回答に対応する詳細な記述を後述する地名・場所名辞書から得る。
- (2) (1) で得られた記述を使って、geocoding<sup>5</sup>を用いて住所やランドマーク名から緯度経度の獲得を行い google マップに表示する。
- (3) 場所を尋ねる質問以外の場合、回答の情報抽出源一つ一つに対し、2.2.2 節で述べた地名補完処理で取得した地名の詳細な記述を得る。
- (4) (3) で得られた記述を使って geocoding を行い、地図上に表示する。

地図上に配置されたマーカーをクリックすると、対応する回答と、その回答が抽出された tweet へのリンクが表示される。意味マップモード同様に、google マップモードもスライダーによって情報抽出源の発信時刻による回答の限定が可能である。

---

<sup>5</sup> <https://developers.google.com/maps/documentation/geocoding/>

## NICT対災害情報分析システムプロトタイプ

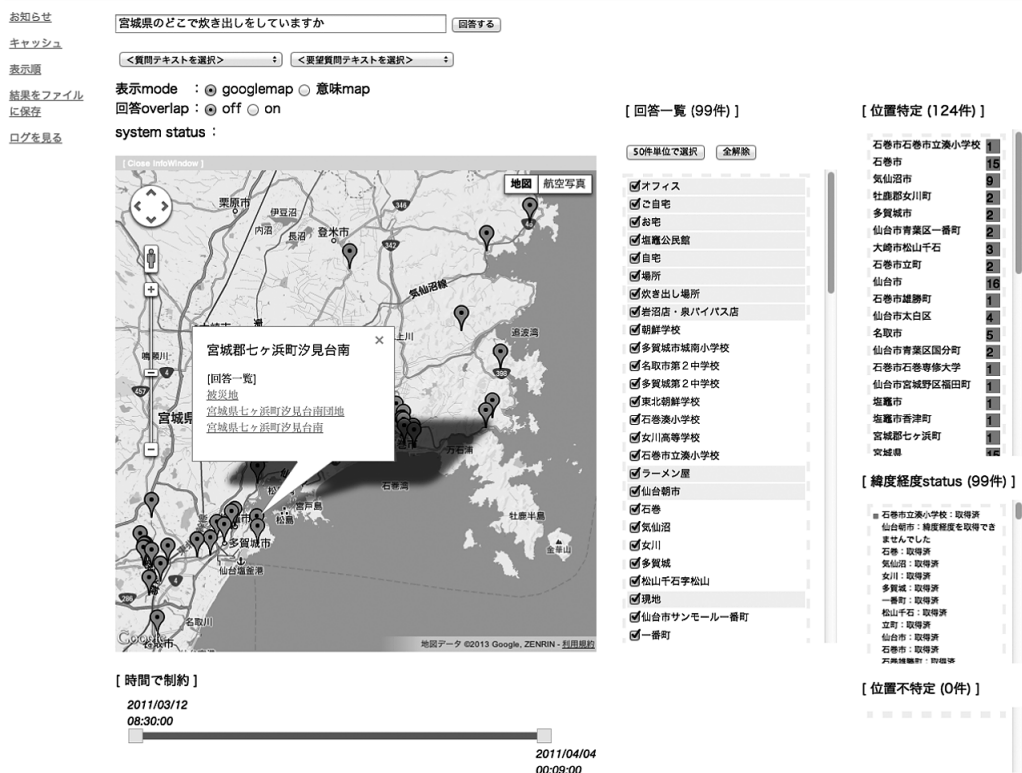


図 4 google マップモードでの実行例

## 3 システムの評価実験

本節では、ここまでで述べた方法を実装したシステムを評価する実験について述べる。システムが実際に運用される場面を想定したシステムの性能を評価することが望ましいが、本論文で提案するシステムは非常に多くのモジュールから構成され、その複雑性や、開発途上にあることを考慮して、システムの基本機能、すなわち質問応答に関して評価を行った。したがって、本論文での実験では入出力モジュールは、直接的にもシステムに組み込まれた形でも評価されていない。

システムを評価するために用いたのは、2011年3月9日から同年4月4日までの tweet データ（約2億2千万 tweet, (株) ホットリンク提供）である。ただし、実験では、災害に関連する345個のキーワードによりフィルターした約5,400万の tweet を用いた。この全 tweet から、システムが回答を取得するためのインデックスとして、約1億2千万エントリを持つ回答インデックス1と、約7億6千万エントリの回答インデックス2（部分パターン用）が生成された。

また、提案システムの評価に加え、次の項目について実験を行った。(1) 含意関係認識における活性・不活性極性の有用性を確認する実験。(2) 固有表現認識器 (NER) の有効性を確認する実験。(3) 教師有り学習を用いた回答のランキングの有効性を確認する実験。このそれぞれについても本節で報告する。

### 3.1 実験の条件

災害時における膨大な情報を整理・分析し、全体的な把握を可能とする本システムでは、入力された質問に対して対象データにおいて目立った回答だけではなく、想定外も含めたロングテール部分に存在する被災者の要望や事実を回答として網羅的に取得する必要がある。そのため、その再現率が重要な評価指標である。

本システムの性能を評価するためにこれまで我々が大規模に作成してきた評価セットを用いる (川田, 大竹, 後藤, 鳥澤 2013)。この評価セットは、6 名で予め作成した質問 300 問の各々について、質問に関連するキーワードでシステムが対象とする tweet を全文検索した結果をランダムに 1,000 件を取得し、その結果から人手で回答を抽出することができた 192 問とその正しい回答 (以下、正答と呼び、その数は 17,524 個である) のセットである。評価セットの正答には質問とは表層的に大きく異なる表現で記載された表現から抽出されたものも多数含まれる。我々が用意した質問は回答が一意に求まるものではなく、ひとつの質問に対して複数の正解が存在する。また、この評価セットは単に質問と正答、つまり名詞句のペアをデータベース化しただけではなく、正答が抽出された tweet も含んでいる。実験では、この評価セットを用いた。再現率は、評価セットに含まれる正答のうちいくつシステムが回答できたかで評価する。当然ながら、評価セットに含まれていないが、正解と判定される回答をシステムが出力することが考えられるが、それを考慮して再現率を計算すると、新たな正解が見つかる度に再現率がかわるため、評価セットに含まれる正答のみ考慮して再現率を求めた。一方、適合率は、システムの回答をランダムサンプルし、正解かどうかを人間が判定して求めた。表 2 に実験に利用した質問の一部を示す。

### 3.2 システムの質問応答性能

評価では、再現率を計算する際に、システムの回答が正答を部分文字列として含んでいるか、システムの回答が正答に部分文字列として含まれているいずれかの場合を正解とした。その結果、再現率 0.519 (9,099/17,524) が得られた。この部分文字列による照合では、正答かシステムの回答が一文字である場合に、多数の回答にマッチし、評価の精度が問題になる可能性があるが、前述したように提案システムは一文字からなる単語を回答として出力しない。また、評価セットの正答で一文字のものは全部で 106 個あったが、システムの出力でそれらにマッチしたものは 67 個であった。これはシステムの回答の 4%程度に相当する。しかし、これらすべてを

表 2 実験に利用した質問例

カテゴリ	例	
インフラ	どこに給水車が来ますか 停電した時の注意点はなんですか	どこでガスが復旧していますか
物資	必要な家電は何ですか 支援物資の受付窓口はどこですか	スポーツドリンク代わりになる物は何ですか
生活	どこでお風呂に入れますか 営業しているお店はどこですか	どこで遺体の火葬をやってもらえますか
ボランティア	ボランティアに適した服装は何ですか どこで復旧作業が行われていますか	介護士のボランティアはどこで募集していますか
支援活動	どこで募金ができますか どこで炊き出しをしていますか	支援が必要なのはどこですか
情報・交通	どこで携帯電話の電波は入りますか 機能している空港はどこですか	どこで道路が寸断していますか
災害状況	津波の高さはどのくらいですか どこで土砂崩れが起きていますか	震災による経済的損失はいくらですか
病気・負傷	はやっている病気は何ですか 不安解消に効くものは何ですか	クラッシュ症候群の注意点はなんですか
原発・汚染	放射能が高いのはどこですか 線量はどのくらいですか	なにが汚染されていますか
安否確認	どこで安否確認ができますか 救援を求めているのはどこですか	どこで身元の確認ができますか
政府・行政	どこに自衛隊がいますか 緊急車両が通れる道はどこですか	自治体の就労支援には何がありますか

回答から除外した場合の再現率は、 $0.519 (= (9,099 - 67) / (17,524 - 106))$ と変わらず、この影響は小さいと考える。また、192 の質問ごとに再現率を求め、その平均をとると 0.428 であった。これは、もともとの正答数が小さい質問において、再現率が 0 となってしまう場合が多い (192 問中 41 問、そのうち回答数が 0 のものは 32 問) ためであり、このことから、逆に質問の正解が得られた場合の再現率は、この数値よりも大きい場合が多いことを期待できる。

適合率に関しては、全回答から質問と回答のペア 250 個をランダムサンプルし 3 名の評価者で正解かどうかを調べ、その多数決により正解を決めた。評価者間の一致度合は Kappa 値 (Fleiss 1971) が 0.507 であった。回答の評価に際しては、回答が抽出された元の tweet が非常に大量の場合があるが、ランダムに選択した最大 3 個の tweet から正解かどうかを判断した。評価の結果、250 問の適合率は、0.608 (152/250) となった。

例えば、構文パターンを利用した質問では、「どこで風評被害が起きていますか」という質問の回答では、「Y で X (= 風評被害) が出ている」「X (= 風評被害) が Y で発生している」「Y で起きている X (= 風評被害)」「X (= 風評被害) が Y で起こる」「X (= 風評被害) が Y で起きている」などのパターンにより回答を取得している。

また、部分パターンを利用した質問では、「なにが汚染していますか」という質問で「Y が汚染されてしまう」「Y が汚染される」「Y の汚染」などのほか、「Y から検出される」「Y からは検出される」などの部分パターンが含意パターンデータベースから取得され利用された。これにより「4 号機、正門、ヘリ」などの tweet に「汚染」を含んでいない回答も得ることができている。

再現率を下げている要因の一つとしては、回答がまったく取得できない質問が 32 問あることがある。これらの多くは、質問文を構成する名詞句が tweet において非常に低頻度であり、手掛かりとして役に立たない場合である。例えば、「専門職ボランティア」、「被災者相談窓口」、「被ばく相談」、「被災者就労支援」などの複合名詞や、「津波肺」「クラッシュ症候群」「誤嚥性肺炎」などの固有名である。これらは、該当する複合名詞や固有名が回答インデックスに存在しないか登録されていても非常に少数であった。対応策としては、「被災者相談窓口」を「被災者の相談窓口」とするなどの複合語の分割が有効であり、さらにサ変名詞を語尾にもつ「被ばく相談」「就学支援」のように複合名詞が「行く」「できる」「実施する」などに係る場合は、「被ばくを相談する」、「就学を支援する」などのより汎用的な表現に変換することが必要である。今後、複合語の構造解析手法などを取り入れ、より幅広い質問にも対応できるようにする予定である。

また、適合率を評価した回答 250 についてより詳細に分析した。これらの回答がどういった処理によって抽出されるかを見るとまず、クラス依存、クラス非依存をふくめて「X が Y で不足している」のように二つの変数を含むパターンによって得られた回答は全体の 6% (15 個) であり、その適合率は 0.933 であった。また、「Y が不足している」のような部分パターンで抽出された回答は 72 % (180 個) を占め、適合率は 0.656 であった。さらに部分パターンの内容語を抽出して得られた回答は 22 % (55 個) であり適合率は 0.364 であった。期待されるように制約の強いパターンで取得されている回答は適合率が高いものの、変数を二つ含む複雑なパターンの適用例はきわめて少なかった。これは「どこが渋滞していますか？」のようなそもそも二つの変数を含むパターンが抽出できない比較的簡単な質問が我々の評価セットに多かったことも理由である。今後「宮城県のどこで渋滞していますか？」のようなより複雑な質問を評価セットに加えると、この制約が強いパターンが適用される割合も増加するものと考ええる。

誤った回答が抽出された要因を見ていくと、もちろん、パターン間の含意の認識誤りも含まれてはいるが、むしろ目立つのは「水は不足していますか?」「水が不足したりして」「水は不足していません」などのように単純な肯定文以外の文から「X が (は) 不足する」のようなパターンが抽出されている場合である。これらの文をムード等の分析ルーチンを導入することによって除くことで最大で 10 % 以上の適合率改善ができると予想している。一方で、「水は不足していますか?」のような質問や要望、「水が不足していたとしたら」のような仮定も、災害時において非常に有用な情報であり、個別に認識することは重要な課題だと考えている。

また、地名補完処理の誤りによって、パターンやその内容語から離れた位置に出現する場所



名が誤って回答として抽出されるケースがあった。これらは今後、省略、照応解析を導入することで改善していく予定である。

### 3.3 部分パターン対のクリーニングの効果

2.3 節で述べた部分パターン間の含意関係のクリーニングが質問応答全体に及ぼす影響について評価を行った。部分パターン間の含意関係とは、例えば「X が崩落する」「X が崩壊する」の間に成立する含意関係である。2.3 節で述べたように、このクリーニングにおいては、活性・不活性極性を用いたクリーニング（活性・不活性クリーニング）、ならびに同一の動詞を含む部分パターン間で助詞のみが異なるものを削除するクリーニング（助詞クリーニング）の二種類を行った。

まず、提案システムの再現率は 0.519、適合率は 0.608 であったが、部分パターン間の含意関係に対して助詞クリーニングのみ適用し、活性・不活性クリーニングを適用しなかった場合の回答を、提案システムと同様に回答 250 サンプル（評価者 3 名による評価）を抽出し、評価したところ、表 3 に示すとおり、再現率 0.524、適合率が 0.536 となった。つまり、再現率は 0.005 とわずかに向上したが、適合率が 0.072 と大きく低下したことになる。さらに、活性・不活性クリーニング、助詞クリーニングの両方を適用しなかったときの性能は、再現率が 0.533、適合率が 0.448 となり、やはり再現率がわずかに向上したものの適合率の大幅な低下が見られた。最終的にいっさいクリーニングを行わなかった場合と提案手法を比べると、再現率が 0.014 程度向上するのに対して、適合率は 0.160 と大幅に低下している。まとめると、部分パターン対のクリーニングは最終的な回答の質において非常に重要であるということが分かった。特に、一見含意関係とは関係の薄い、活性・不活性という意味極性がそのクリーニングにおいて重要な役割を果たすことが確認できた。

### 3.4 固有表現認識器の効果

本研究での提案システムは地名補完モジュールに NER を使用しなかったが、それは以下の実験結果により、NER の有用性が本システムにおいて認められなかったからである。

まず、IREX 固有表現コーパス (Sekine and Isahara 2000) において LOCATION タグのみを残し、これを NER 学習データ 1 とした。次に、Twitter API を使用して、実験で用いる tweet と

表 3 部分パターン対のクリーニングの効果

	再現率	適合率
提案システム	0.519	0.608
活性・不活性クリーニング未使用	0.524	0.536
活性・不活性及び助詞クリーニング未使用	0.533	0.448

は異なる期間の tweet 22 万 5 千件を取得し、これに対し、災害関連のキーワード 345 個のいずれかを含む 11 万 tweet に対して学習データ 1 から作成した既存の NER を適用し、LOCATION タグを付与した。この結果のうち 4 万文を人手で修正し、これを NER 学習データ 2 とした。これらの NER 学習データ 1 ならびに 2 をあわせて NER 構成用学習データとし、CRF++<sup>6</sup>を用いて形態素単位の NER を構成した。素性テンプレートは CRF++ パッケージのサンプルとして含まれているものをそのまま利用した。

この NER を評価するために、我々が対象としている 5,400 万の tweet から 1,000 tweet (3,017 文) をランダムサンプルし、構成した NER を適用した。その結果を人手で修正し、評価用テストセットを作成した。この評価用テストセットの形態素数は約 33,000 であり、LOCATION とされる名詞句は、521 (866 形態素) 存在する。これを用いて構成した NER を評価したところ適合率は 0.930、再現率は 0.839 であった。

次に我々の質問応答システムで、地名補完モジュールにおける処理対象の特定に NER を組み入れた場合と、形態素単位の文字列によって直接辞書引きすることで特定する場合との違いがシステム全体の質問応答性能に与える影響を調べた。実験に使用したのは、部分パターン対のクリーニングを行う前のシステムであるが、NER の効果を調べるには問題がないと考える。表 4 に示すとおり、実験結果は、NER を用いない場合が再現率 0.533、適合率 0.448 であり、NER を用いた場合には再現率 0.516、適合率 0.392 と再現率、適合率ともに低下した。この結果から、あるエンティティが地名・場所名辞書に存在しているにもかかわらず、NER がそれを特定できなかった場合や、逆に NER が地名補完モジュールでの処理対象を特定できても地名・場所名辞書に登録されていない場合などがあり、地名・場所名辞書を直接辞書引きしたほうが、より高い性能を発揮できたと考える。

より具体的に、NER で特定されたものがどれだけ地名・場所名辞書を用いて地名補完処理されたかを見てみると次のようになった。NER はテストセットに 521 あるエンティティのうち、437 (再現率 0.839) 相当を正しく特定できているが、このうち、地名補完処理の対象となったのは、わずか 157 個である。この数字が小さい理由は、現在の地名補完処理はシステムの持つ地名・場所名辞書にあるエントリしか処理対象としないからであり、さらには NER の認識結果と地名・場所名辞書との食い違いが大きいからである。一方、地名補完モジュールにて行って

表 4 固有表現認識器 (NER) の効果

	再現率	適合率
固有表現認識 (NER) 未使用	0.533	0.448
固有表現認識 (NER) 使用	0.516	0.392

<sup>6</sup> <http://crfpp.googlecode.com/svn/trunk/doc/index.html>

いる処理では、214 個の地名・場所名を特定し、地名補完処理がなされた。もちろん、この地名補完処理がなされた地名・場所名には誤ったものも多数含まれていよう。もともと NER を導入した動機は、NER によって一般名詞や人名等を地名として誤認識することを防げるかもしれないということであった。つまり、地名補完処理対象認識の適合率の向上をねらったということである。おそらく、地名・場所名の誤認識が NER によって防がれたケースもあったものと推測されるが、そもそも地名補完処理が起動されないことのデメリットの方が大きく、最終的な質問回答の性能が低下したものと考ええる。

もちろん、今後 NER の認識結果を地名・場所名辞書に追加していくことによって、性能向上を見ることは可能かもしれない。しかしながら、そこで障害となるのはエンティティの基準と、地名補完処理において処理対象とするエンティティ、すなわち地名・場所名辞書のエントリの認定基準とが異なっていることである。例えば、NER の認識結果には外国の地名などあきらかに本タスクでは不要と思われるものも多数存在するし、複合名詞中、例えば「富士スピードウェイ」の「富士」が地名として認識されるといった問題も存在する。また、地名・場所名辞書では地名間の包含関係が情報として含まれているが、NER の認識結果にはそうしたものは含まれない。これらの問題をどう解決していくかが、今後の課題の一つとなる。

まとめると、風間ら (風間 他 2012) の報告と同様に提案システムにおける NER の効果は確認出来なかった。これをうけて、我々の提案システムでは NER を使用していない。この理由は、現状の地名補完処理では、固有表現特定後に地名・場所名辞書にて詳細な地名情報を取得する必要がある、この辞書の網羅性等が性能に影響するためである。さらには、地元でだけ用いられる通称など考慮しなければならない点もあり、これらの問題点をいかに低コストで解決していくかも重要な点であると考えている。今後、自治体などの協力を得て、そうした通称や未登録の避難所をリストアップしていくなどの作業も必要であろう。したがって、システムの性能を向上させるためには、NER の認定基準と本タスクで必要とされる地名・場所名の認定基準との擦り合わせ、さらには地名・場所名辞書との整合性をとる自明でない作業が必要となる。

### 3.5 回答のランキング

本論文におけるシステムでは、ロングテールに存在する回答についてもすべて出力するという目的から、再現率を重視し、今まで述べてきた手法で発見できたすべての回答を出力している。一方で、自明な拡張は回答にランキングメカニズムを導入し、さらなる拡張を図ることである。本来、再現率を重視しつつ、ランキングを導入し、提案手法よりも高い性能を達成するためには、提案手法よりも公汎な回答を出力し、ランキングに基づいて回答の足切りを行うべきであるが、現状はそこまでの実験は行えていない。代わりに、提案システムが出力する回答全部を教師あり学習に基づいてランキングした結果について報告する。

今回行った実験で使ったランキング手法は、回答とパターンに関する素性をもとに学習した

SVM のスコアによりランキングを行うものである。表 5 に, SVM の学習に利用した素性を示す。

まず, パターンの属性に基づく素性として, 質問構文パターン, クラス依存パターン, クラス非依存パターン, 部分パターンからのいずれのパターンで回答が得られたか, あるいは部分パターンと部分パターンの内容語によるキーワード検索を用いたかを示す 2 値の素性を用いる。これに加え, クラス依存パターン, クラス非依存パターン, 部分パターンの各スコアを用いる。ある回答が複数の異なるパターンから得られた場合には, その全パターン数, パターンが回答を連体修飾していないかどうか, 全パターン数と回答を連体修飾していないパターンとの比率を利用する。また回答を抽出した含意パターンや部分パターンが, 質問構文パターンと共通の漢字を持つかどうかにも利用する。

回答の属性に基づく素性では, まず, 様々なパターンから得られた同じ回答の個数, その文

表 5 回答のランキングに使用する素性一覧

素性の種類	内容
パターン	質問構文パターンにより回答が得られたかどうか
パターン	クラス非依存パターンにより回答が得られたかどうか
パターン	クラス依存パターンにより回答が得られたかどうか
パターン	部分パターンにより回答が得られたかどうか
パターン	部分パターンでのキーワード検索により回答が得られたかどうか
パターン	部分パターンの内容語でのキーワード検索により回答が得られたかどうか
パターン	クラス依存パターンのスコア
パターン	クラス非依存パターンのスコア
パターン	部分パターンのスコア
パターン	回答を取得したパターン数
パターン	連体修飾型でないパターンの有無
パターン	回答取得したパターン数と連体修飾型でないパターンとの比率
パターン	質問構文パターンと同じ漢字を含むパターンの有無
回答	複数のパターンから得られた同じ回答の個数
回答	回答の形態素数
回答	回答の文字数
回答	回答の意味クラス
回答	未特定の意味クラスかどうか
回答	部分文字列によるクラス特定を利用したかどうか
回答	構文パターンと 2 つの名詞句の意味クラスの PMI
回答	質問構文パターンと質問文中の名詞に基づく回答の意味クラスの尤度
回答	疑問代名詞タイプ
回答	回答が疑問代名詞の対応するクラスに属するかどうか
回答	回答と主題語との分布類似度
回答	回答が主題語の下位概念かどうか
回答	回答の末尾が主題語となるかどうか

字数及び形態素数を用いる。次に、回答の意味的な情報として、回答の意味クラス、その意味クラスを特定する際に部分文字列を用いたか、回答のクラスが未特定かどうかの2値の素性を用いる。また、回答を獲得したパターントリプルの構文パターンと2つの名詞句の意味クラスのPMI（相互情報量：Point-wise Mutual Information）、質問構文パターンと質問文中の名詞に基づく回答の意味クラスの尤度 (De Saeger et al. 2009) を利用する。また質問文から得られる疑問代名詞と主題語を利用した素性として、疑問代名詞タイプ、回答が疑問代名詞の対応するクラスに属するかどうか、回答と主題語との分布類似度、回答が主題語の下位概念となるかどうか、回答の末尾に主題語を含むかどうかを用いる。

上記の素性を用いて、線形、多項式（二次）、放射基底関数（RBF, 比例定数 1）の各カーネルを用いて SVM の学習を行い、いずれのカーネル関数を用いるべきか検討した。学習データは、災害に関連の深い質問 60 問（これまでの評価で利用した質問とは別である）と、システムが出力した回答のペア合計 5,044 個に対して正解／不正解のラベルを付与したデータである。なお、このデータは提案システムの古いバージョン、つまり、場所名フィルターや部分パターン含意データベースのクリーニングを行っていないシステムの出力を含んでおり、現在の提案システムでは出力できない回答も含まれている。10 分割交叉検定の結果、線形カーネルで F 値 0.642（適合率 0.681, 再現率 0.607）、多項式カーネルで F 値 0.631（適合率 0.626, 再現率 0.634）、RBF カーネルで F 値 0.529（適合率 0.719, 再現率 0.419）が得られた。本システムでは F 値が最も高かった線形カーネルにより学習された分類器の出力するスコアを利用することを検討した。

3.2 節の実験にて利用した 250 個の回答サンプル（適合率 0.608）を以上の分類器のスコアでランキングした結果が図 5 である。グラフの再現率は提案システムの出力すべてを SVM のスコアにしきい値をもうけて足切りを行い、足切りを生き延びた回答集合を 17,254 件の正解データ

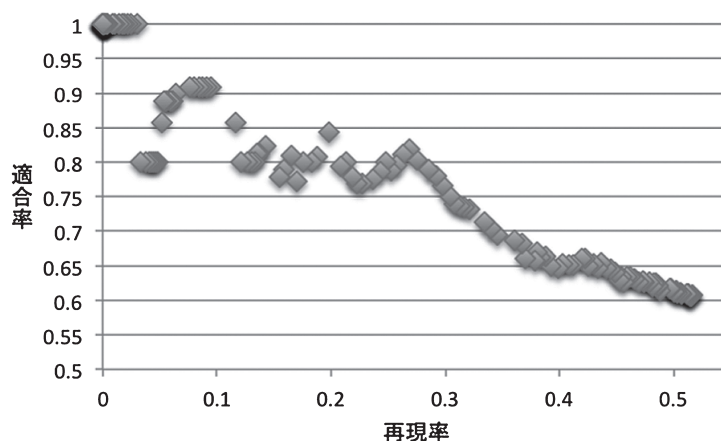


図 5 回答のランキング結果

に照らし合わせて計算されたものである。これによると、再現率が0.1前後のところでは適合率が0.90前後であり、きわめて高いものとなっている一方、システムの全回答の再現率0.508に近いところ、例えば、再現率0.4前後のところでは提案手法の適合率に比して、わずかな適合率の向上（0.05前後）しか見られず、また、もうすこし離れたデータポイント（例えば、再現率0.3前後のデータポイント）までの適合率の改善具合もきわめてなだらかである。

この評価はあくまで現状のシステムの出力結果のみをランキングしているため、確定的なこととは言えないが、前述したように学習データは現在のシステムが出力できない回答に関するものも含まれていないことも考え合わせると、仮に現在のシステムをより大量の回答を出力するように改変し、ランキングによる足切りをおこなったとしても、例えば、再現率0.5前後の部分での適合率向上はきわめて小さなものになる可能性が高いと考えられる。これは再現率を重視するという我々の立場とは相容れないものであり、提案システムにはランキング手法は導入しなかった。

一つ今後システムを改善できる可能性があるとするれば、今後さらに学習データを増やしていくことが重要であるが、現在でも約5,000件という少なくない量の学習データを利用していること、また、次の災害はおそらく東日本大震災とは大きく異なることが予想され、東日本大震災に特化した学習データを作ることは望ましくないと考えられることから、少なくともランキング手法の導入については慎重に検討する必要があると考えている。実際に大規模な災害が発生した後、アノテーションをクラウドソーシングなどで行い、質問応答の精度を高めるといったシナリオは魅力的に見えるかもしれない。しかし、そうしたシナリオを実現するためには、NERの場合と同様にシステム全体としての最適化の枠組みなどが必要だということかもしれない。これも慎重に検討する必要があると考えている。

#### 4 さらに行き届いた被災情報の活用を目指して

本システムはインターネット経由で得られる情報を収集・分析し、ユーザからの質問に備える。図6は、本システムを災害時にどのように運用するかを示したイメージ図である。各種救援団体、例えば、炊き出しを行うボランティア団体などは、自らの炊き出し実施場所を決めるためにどこで炊き出しを行っているかをシステムに質問し、そのすべての回答を地図上に表示することで、炊き出しが行われていないエリアを確認できる。一方、被災者など個人レベルで本システムを利用する場合には、自分の周辺の状況を把握し、意思決定の助けとするような使い方や、また把握した状況に基づき、自らの周辺状況や救援要請を発信するなどの使い方を考えている。このように、本システムは災害時において、ソーシャルメディア等に溢れる情報を整理し、救援団体や、自治体、被災者らに対して被災状況の全体的把握を容易にする情報をわかりやすく提示することで、被災者の救援・支援に有効である。

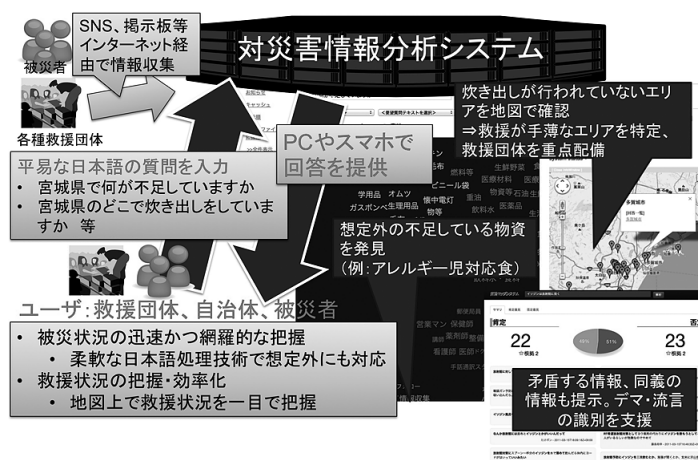


図 6 災害時における提案システム運用イメージ

一方で、災害時には、通信状況等様々な制約から、回答のすべてを確認することが困難な状況も考えられる。そこで、重要と考えられる回答の一部を表示するために、結果をランキングできることが望ましいが、質問に対する一般的な回答の適切さのみならず、過去5分以内に挙げられた情報を求める場合のように情報の新鮮さを重視する場合や、回答の利用目的（見落としているかもしれないものにはやく気づきたい）などによっていくつかの基準が考えられる。時刻による限定は、現在機能として有しているが、ランキングの基準とあわせて今後利用者にとってさらに使いやすくすべきである。

さらに、インターネット経由でつぎつぎに情報（テキスト）が流れ込んでくる状況においては、システムが大規模コーパスから獲得して利用している知識、例えば、意味クラス辞書や、含意パターンデータベースを拡張可能かもしれない。しかしながら、これらのデータベースは、一度、大規模なコーパスから獲得してしまえば、大部分のものは長く使えるものである。特に含意パターンデータベースは、名詞句が変数となっており、その経時的変化は非常にゆるやかであると考えられる。災害発生後にそれまで使っていたパターンとは全く異なるパターンで情報発信することは考えにくい。したがって、事前に大規模なコーパスから獲得した知識を用いていることによって損なわれる有用性は非常に限定されると考える。もちろん、オンライン学習等によって常時知識が更新されつづけるようシステムに拡張すべきであることは言うまでもない。

本論文の冒頭で示唆したように、今回の震災時には被災者からの tweet が必ずしも救援者へ届いていないという問題があったようである。本システムは、被災地からの情報を全体的俯瞰的に把握することを可能とする。しかしながら、一度質問した内容でも、対応する情報は被災地の各地から質問後も不定期に投稿される可能性が高く、その情報は常に更新される。比較的落ち着いた時期になれば、定期的に分析システムを利用すればよいが、災害時に様々な対応が

必要な自治体などの支援者側は思うようには反応することができないことが予想される。また、情報発信を行っている被災者サイドでも発信した情報が適切な救援者に届いているか否かが不明な状況では、例えばさらに遠くへ避難するか、それとも救助を待つかといった切迫した判断を行えないといった問題が生じる。

そこで、我々は、図7に示すように、本システムの回答インデックス作成モジュールを拡張し、予め救援者がシステムに登録した質問に対しては、以後の tweet や指定した BBS、掲示板に情報が発信された場合に、システムがその内容が登録済みの質問の回答となるかをリアルタイムで判断し、救援者サイドの情報のアップデートを行うとともに、情報提供者にも、質問を登録した救援者にその情報が届いたことが通知される仕組みを開発している。この処理により、図8のように、情報提供者、被災者は自らの発信した情報が救援者に届いたことがわかりその後の意思決定が容易になるとともに、救援者側は欲している情報をリアルタイムで定常的に取得

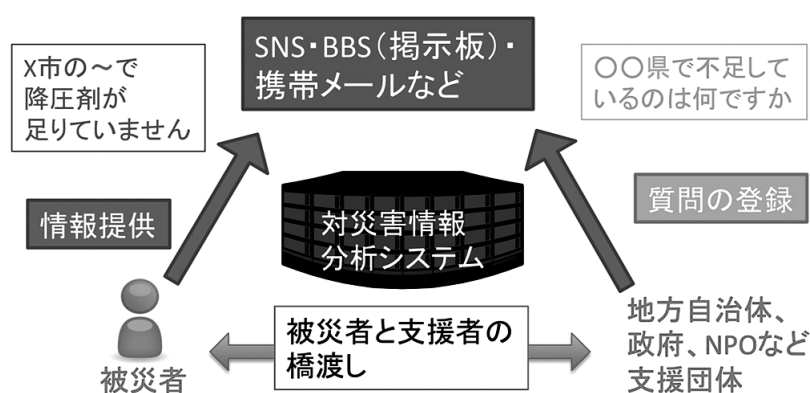


図7 被災者と救援者の双方向のコミュニケーション

## NICT 対災害情報分析掲示板 プロトタイプ

[掲示板に戻る](#) [全部](#) [最新50](#)

宮城県で不足しているもの

12 : demo\_user 2013/04/02(火) 17:48:27

石巻市で防寒着が不足しています。

13 : 対災害質問応答システムボット 2013/04/02(火) 17:48:27

>>12

NictNpoさん の質問「石巻市では何が不足していますか」の回答として認識されました。  
NictNpoさん のPCで表示されているものと思われます。

図8 掲示板上的動作例



することができ、支援のスピードアップにつながると考えられる。こうした一連の操作は、一言で言えば、現状のマイクロブログ、SNS、掲示板等のいわば一方通行の情報提供から、被災者サイドと支援団体等救援者サイドの双方向のコミュニケーションを担保することとも言え、こうした操作によってよりスピーディかつ適切な救援、避難等が実現できるのではないかと考えている。

こうした処理は、これまでに説明した質問応答の処理の方向を大幅に変更することなく実現できる。通常の質問応答処理では、パターントリプルもしくは部分パターンの形式でインデックスに登録された tweet からの情報を、質問から取得したパターントリプル等を含むクエリにより検索するが、ここでは、あらかじめ登録された質問に対して、含意パターンなどの獲得を事前にやっておき、含意パターンも含むようなパターントリプル等をキー、質問を値とする別種のいわば質問のインデックスを作成しておく。

例えば、「宮城県で不足しているのは何ですか?」といった質問が登録されているとするならば、「X で足りない Y」といった含意パターンや、「宮城県」といった名詞句を含むパターントリプルをキーとし、「宮城県で不足しているのは何ですか」という質問を値とするような質問のインデックスが作成される。掲示板等の記事や tweet が新規にシステムに渡されると、将来問われる質問にそなえてこれまでに説明してきた回答インデックスが作成されるが、その際、生成されるパターントリプルをキーとして、過去に登録された質問のインデックスを検索する。もしこの質問のインデックスの検索がヒットすれば、値となっている登録済み質問の回答をアップデートするとともに、対応する新規の tweet、記事等の作者に対して、登録された質問への回答として提供された情報が認識されたことを通知する。現状は、こうした枠組みをサーバー一台の上で動作させることができており、今後、大規模な計算機クラスタ上等で想定されるような大量の情報がやってきたときでもリアルタイムの処理が可能なシステムを開発していく予定である。

本システムのもう一つの応用としては、ソーシャルメディア上で流通している様々なデマの早期発見とエキスパートによる反論を支援するものが考えられる。例えば、図 9 で「放射能に効くのは何ですか」という質問に対してのシステム出力を示す、「イソジン」、「わかめ」、「活性炭」など、デマと思われるものが大量に含まれる。このような質問も事前にシステムに登録することで、信頼性が低い情報あるいは有害情報が爆発的に拡散される前に、書き込まれた時点で認識され、デマが大量に拡散する以前にエキスパートによってデマを打ち消す情報をスピーディに発信することが可能となると考えられる。

また、本システムが提示する回答にはそもそも大量のデマが含まれている可能性があるが、我々は本システムを東北大学で開発されている言論マップ (水野, Nicoles, 渡邊, 村上, 松吉, 大木, 乾, 松本 2011) と組み合わせることで回答を閲覧したユーザが回答のデマ性についてより適切な判断を下すことができるようになって考えており、実際に言論マップとの統合を計画して

## NICT対災害情報分析システムプロトタイプ



図 9 「放射能に効くのは何ですか」という質問に対するシステム出力

いる。現在の言論マップでは、例えば「イソジンは放射能に効く」という情報に対して、それを肯定している情報と否定している情報をソーシャルメディア上の情報から発見して提示することが可能である。こうした肯定的情報、否定的情報は通常のソーシャルメディアの閲覧環境では簡単に見つけることは難しいが、本システムに言論マップを組み合わせることで、回答には常に肯定的情報、否定的情報をあわせて表示することが可能となり、ユーザは疑わしい情報に関しては、こうした情報を参考にしつつその真偽を判断する材料とすることができる。

## 5 関連研究

近年では、検索エンジンや質問応答システムなど、情報へのアクセス手段の進歩が目覚ましい。例えば、質問応答システムとしてはIBM社のWatson (Ferrucci, Brown, Chu-Carroll, Fan, Gondek, Kalyanpur, Lally, Murdock, Nyberg, Prager, Schlaefel, and Welty 2010) がクイズ番組の人間のチャンピオンに圧勝し一躍有名になった。Watsonは、Wikipediaを含む辞書、辞典や台本などJeopardyというクイズ番組の分野に関連する確かな知識を予め選別し、データベース化している。少なくとも我々の知る限り、情報のリアルタイム更新については想定していない

ため、逐次更新される災害時の情報などには対応していない。また、災害時に必要なのは、多数の情報を俯瞰的に閲覧することであるが、すくなくとも Jeopardy において質問はその回答が一意に定まるものに限られており、Watson が現状すぐに災害時の情報などに適用できるかどうかは不明確である。

また、日本においても、「しゃべってコンシェル」(吉村 2012) と呼ばれる携帯電話のサービスが注目を浴びている。このシステムは、携帯電話を用いて質問応答を行うものであり、Web の更新データに対応している。このサービスでの質問応答では、「ハリーポッターの監督は誰」のようなある対象物（ハリーポッター）の属性（監督）を聞くタイプの質問は、回答が一意に定まる知識について Watson と同じように予め知識のデータベース化を行っている。また、天気やニュースなどについては専用のサービスの結果を返し、それ以外の質問は、キーワード検索と固有表現による質問応答手法を利用している。それでも見つからない場合はキーワードによる検索結果を出力する。

我々のシステムは、上記のシステムとは異なり、これまでの含意獲得の研究をもとに、質問文からの含意パターンや部分パターンを取り出し、そのパターンを元に回答を求めている。そのため、質問文から何らかのパターンが獲得できれば、高い精度で回答が可能である。また、固有表現でない一般名詞が回答の場合や、これまでの固有表現 (Sekine and Isahara 2000; Sekine 2008) では対象としていない表現についても、回答を出力できる。固有表現は、特定の質問に対しては重要な要素であることは間違いないため、今後、回答のランキングに固有表現に関する素性を取り入れて行く予定である。また、災害時の質問にも、一意に定まる質問がされる可能性はあるため、Wikipedia などの知識を利用した手法 (Buscaldi and Rosso 2006) もシステムも取り入れて行く。

## 6 おわりに

本論文では、想定外のものもふくめて、災害時の情報を俯瞰的に把握するために開発した、質問応答に基づく情報分析システムについて述べ、また、東日本大震災時の tweet データを利用した性能評価について報告した。さらに、本システムを拡張することによって、比較的实现容易な形で、リアルタイムで被災者と救援者が双方向のコミュニケーションを行うことを可能とし、より効率的な救援活動やより適切な避難行動等を支援する枠組みについても提案した。また、東北大学で開発されている言論マップ技術との統合や、リアルタイムでの回答の更新によって、東日本大震災時に問題となったデマに対処する枠組みも提案した。

様々な質問に対して回答を提示できるようにするために、本システムでは、質問応答処理において構文パターンの言い換えに基づく質問文の拡張を行い、さらに場所や地名の補完処理を加えることで、幅広い質問に対応した。また、得られた回答を意味クラスごとにまとめるイン

ターフェースと回答を地図上に表示するインターフェースを用意することで被災地の状況や救援状況の俯瞰的把握を可能とした。人手で作成した質問を基にした評価実験では、複合語処理の問題や要望、疑問、仮定を含む tweet の特定の必要性が明らかになった。必要な情報を必要な人に行き渡らせるためには、たとえその回答を必要としている人が一人であっても、回答を提示することが望ましい。こうした点に鑑み、ロングテール部分に存在する被災者の要望にも応えることができる情報分析システムの構築を今後も進めていく予定である。

より具体的には、現在5万件以上からなる災害に関連の深い含意パターンの学習データを人手で構築しつつあり、さらに、その他の言語資源も構築中である。今後、こうしたリソースを活用しつつ、また、新規のアルゴリズムを導入することによって性能向上を図っていく予定である。

## 謝 辞

本研究で利用しているデータは、株式会社ホットリンク様よりご提供頂きました。ここに記して感謝致します。

## 参考文献

- Bloom, B. (1970). "Space/time Trade-offs in Hash Coding with Allowable Errors." *Communications of the ACM*, **13** (7), pp. 422–426.
- Buscaldi, D. and Rosso, P. (2006). "Mining Knowledge from Wikipedia for the Question Answering Task." In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pp. 727–730.
- Cheng, Z., Caverlee, J., and Lee, K. (2010). "You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users." In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 759–768.
- De Saeger, S., Torisawa, K., Kazama, J., Kuroda, K., and Murata, M. (2009). "Large Scale Relation Acquisition using Class Dependent Patterns." In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, pp. 764–769.
- Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., Lally, A., Murdock, J. W., Nyberg, E., Prager, J., Schlaefer, N., and Welty, C. (2010). "Building Watson: An Overview of the DeepQA Project." *AI Magazine*, **31** (3), pp. 59–79.
- Fleiss, J. (1971). "Measuring Nominal Scale Agreement among Many Raters." *Psychological Bulletin*, **76** (5), pp. 378–382.

- Harris, Z. (1954). “Distributional Structure.” *Word*, **10** (23), pp. 142–146.
- Hashimoto, C., Torisawa, K., De Saeger, S., Oh, J., and Kazama, J. (2012). “Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web.” In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 619–630.
- 川田拓也, 大竹清敬, 後藤淳, 鳥澤健太郎 (2013). 災害対応質問応答システム構築に向けた質問・回答コーパスの構築. 言語処理学会第 19 回年次大会発表論文集, pp. 480–483.
- 風間淳一, De Saeger, S., 鳥澤健太郎, 後藤淳, Varga, I. (2012). 災害時情報への質問応答システムの適用の試み. 言語処理学会第 18 回年次大会講演論文集, pp. 903–906.
- Kazama, J. and Torisawa, K. (2008). “Inducing Gazetteers for Named Entity Recognition by Large-Scale Clustering of Dependency Relations.” In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-08: HLT)*, pp. 407–415.
- Kloetzer, J., De Saeger, S., Torisawa, K., Sano, M., Goto, J., Hashimoto, C., and Oh, J. (2012). “Supervised Recognition of Entailment between Patterns.” 言語処理学会第 18 回年次大会講演論文集, pp. 431–434.
- Liu, X., Wei, F., Zhang, S., and Zhou, M. (2013). “Named Entity Recognition for Tweets.” *ACM Transactions on Intelligent Systems and Technology*, **4** (1), pp. 1–15.
- 水野淳太, Nicoles, E., 渡邊陽太郎, 村上浩司, 松吉俊, 大木環美, 乾健太郎, 松本裕治 (2011). 言論マップ生成技術の現状と課題. 言語処理学会第 17 回年次大会講演論文集, pp. 49–52.
- Neubig, G., Matsubayashi, Y., Hagiwara, M., and Murakami, K. (2011). “Safety Information Mining—What Can NLP Do in a Disaster—.” In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP)*, pp. 965–973.
- Ritter, A., Clark, S., Mausam, and Etzioni, O. (2011). “Named Entity Recognition in Tweets: An Experimental Study.” In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1524–1534.
- Sekine, S. (2008). “Extended Named Entity Ontology with Attribute Information.” In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*.
- Sekine, S. and Isahara, H. (2000). “IREX: IR and IE Evaluation Project in Japanese.” In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pp. 1475–1480.
- Yamada, I., Torisawa, K., Kazama, J., Kuroda, K., Murata, M., De Saeger, S., Bond, F., and Sumida, A. (2009). “Hypernym Discovery Based on Distributional Similarity and Hierarchical Structures.” In *Proceedings of the 2009 Conference on Empirical Methods in Natural*

*Language Processing (EMNLP)*, pp. 1172–1181.

吉村健 (2012). しゃべってコンシェルと言語処理. *IPSJ SIG Technical Report Vol. 2012-SLP-93* (4), pp. 1–6.

## 略歴

**後藤 淳**：1993 年徳島大学大学院工学研究科修士課程修了。同年日本放送協会入局。2011 年より独立行政法人情報通信研究機構専門研究員。現在、総合研究大学院大学博士課程在学。

**大竹 清敬**：2001 年豊橋技術科学大学大学院博士後期課程修了。博士（工学）。同年より株式会社 ATR 音声言語コミュニケーション研究所。2006 年より独立行政法人情報通信研究機構。音声言語処理、自然言語処理の研究に従事。

**Stijn De Saeger**：2006 年北陸先端科学技術大学院大学博士課程修了。博士（知識科学）。独立行政法人情報通信研究機構専攻研究員を経て、現在同機構主任研究員。知識の自動獲得の研究に従事。言語処理学会第 16 回年次大会優秀発表賞等受賞。

**橋本 力**：2005 年京都大学研究員、2007 年山形大学助教、2009 年独立行政法人情報通信研究機構専攻研究員。現在、同機構主任研究員。博士（言語科学、情報学）。情報処理学会論文賞、言語処理学会論文賞、同学会優秀発表賞等受賞。

**Julien Kloetzer**：2006 年パリ第 6 大学卒業、2010 年北陸先端科学技術大学院大学博士課程修了。博士（情報科学）。2011 年独立行政法人情報通信研究機構入所。現在、同機構情報分析研究室研究員。

**川田 拓也**：2003 年国際基督教大学教養学部卒業。2010 年京都大学大学院文学研究科博士後期課程修了。博士（文学）。現在、独立行政法人情報通信研究機構情報分析研究室研究員。言語資源の設計と構築に従事。

**鳥澤健太郎**：1995 年東京大学大学院理学系研究科中退。同年同専攻助手。北陸先端科学技術大学院大学助教授を経て、現在、独立行政法人情報通信研究機構・情報分析研究室室長及び情報配信基盤研究室室長。博士（理学）。日本学術振興会賞など受賞。

(2012 年 12 月 14 日 受付)

(2013 年 3 月 6 日 再受付)

(2013 年 3 月 31 日 採録)